Rule WLM220: Service Class was delayed because of resource capping

Finding: CPExpert has determined that resource capping was a major cause of the service class not achieving its performance goal.

Impact: The impact of this finding depends upon the amount of resource capping delay experienced by the service class. A high percent of resource capping delay means HIGH IMPACT while a low percent of resource capping means LOW IMPACT. See the output associated with the rule which caused this rule to be invoked (Rule WLM101 to Rule WLM103, depending upon the type of service class and performance goal).

Logic flow: The following rules cause this rule to be invoked:

Rule WLM101:	Service Class did not achieve average response goal
Rule WLM102:	Service Class did not achieve percentile response goal
Rule WLM103:	Service Class did not achieve execution velocity goal

Discussion: Resource capping is a way of controlling the distribution of CPU service to one or more service classes. Resource capping is implemented by defining "resource groups" to the Workload Manager. A resource group is simply a named set of two values: a minimum CPU service specification and a maximum CPU service specification. The specifications are in terms of **unweighted CPU service units** (that is, the CPU service coefficients are not applied to TCB nor SRB raw CPU service units).

The Workload Manager will attempt to provide the minimum CPU service to the resource group and will restrict the resource from using more than the maximum CPU service.

Service classes are associated with resource groups; however, a particular service class can be associated with only one resource group[1].

It normally is not advisable to use resource groups. IBM provides the facility solely for special cases, and IBM does not contemplate resource groups being normally used. Resource group specifications are "preemptive" in nature, in that the Workload Manager attempts to honor resource group specifications before considering other service

[1]Please see Section 4 (Chapter 1.6) for a discussion of resource groups and how the Workload Manager implements the resource group specifications.

specifications. Consequently, **resource group specifications could nullify the rest of the Workload Manager's algorithms**.

When the maximum CPU service specified in the resource group has been used, the Workload Manager marks "non-dispatchable" the TCBs and SRBs associated with the service classes assigned to the resource group. This is the situation addressed by Rule WLM220.

As the System Resources Manager takes its samples of the state of address spaces, it examines whether a dispatchable unit (TCB or SRB) is marked non-dispatchable because of a resource group maximum. Samples reflecting the resource group maximum are recorded by RMF in the SMF Type 72 delay samples, as CPU Capping Delay (R723CCCA).

CPExpert computes the percent of CPU Capping Delay for the service class, as a function of the overall execution of transactions executing in the service class. CPExpert produces Rule WLM220 if the percent of CPU Capping Delay for the service class is greater than the significance value specified in the **WLMSIG** guidance variable in USOURCE(WLMGUIDE).

With Rule WLM220, CPExpert provides the total number of ending transactions in the RMF measurement interval, the total CPU service units consumed by the transactions, the average CPU service units per transaction, and the average percent resource capping delay to transactions active in the service class.

The following example illustrates the output from Rule WLM220:

```
RULE WLM220: SERVICE CLASS WAS DELAYED BECAUSE OF RESOURCE CAPPING

   Service Class BATCH (Period 1) was delayed waiting for CPU
   resource capping.  This means that a TCB or SRB in the Service Class
   was marked non-dispatchable because the Resource Group maximum was
   being enforced.  Service Class BATCH (Period 1) was assigned
   Resource Group BATCHCAP, which specified a maximum of 500 CPU
   service units per second.  This situation applies to the following
   measurement intervals:

                                                                  AVG %
                          TOTAL     TOTAL CPU      AVERAGE CPU   CAPPING
   MEASUREMENT INTERVAL   TRANS   SERVICE UNITS   SERVICE UNITS   DELAY
   15:00-15:16,01MAR1994    658      452,664           876         22.1
```

Suggestion: As mentioned above, resource groups are intended for very special situations. In most environments, it is far better to allow the Workload Manager to manage system resources to meet the performance goals specified for various service classes. Using resource groups takes control away from the Workload Manager.

Further, specifying maximum CPU service units may result in unused CPU capacity if there are no other service classes ready to use the CPU service.

CPExpert suggests that you consider the following alternatives:

- While there may be unusual situations in which control must be removed from the Workload Manager, please consider whether you have such an unusual situation. If you do not have an unusual situation, you may wish to remove the resource group specification from the service class definition.

- Alternatively, you should review the performance goal specified for the service class identified by Rule WLM220. CPExpert performs "delay analysis" only on service classes which fail to achieve their performance goal. Consequently, the service class identified by Rule WLM220 had failed to achieve its performance goal.

  The performance goal may be incompatible with the resource group Capacity Maximum, and you may wish to either increase the performance goal (for response goals) or decrease the performance goal (for execution velocity goals).

- Alternatively, you should review the CPU usage report produced by CPExpert at the end of the normal rule listing. Compare the CPU time used by the service class identified by Rule WLM220 with the CPU time used by other service classes. Pay particular attention to CPU time used by any service classes at the same or lower importance, to see whether these service classes should receive the CPU service indicated.

- Alternatively, you may wish to increase the Capacity Maximum specified for the resource group. Since applications executing in the service class are being delayed because of CPU capping, you may remove or decrease the delay by increasing the Capacity Maximum for the resource group.

- Alternatively, you may wish to review the applications executing in the service class identified by Rule WLM220, to determine whether the application code can be optimized so that less CPU time is required.

- If none of the above alternatives apply and if Rule WLM220 continually is produced for the service class, you may wish to exclude the service class from CPExpert's analysis[2]. There is little point in having findings produced which cannot be acted upon.

[2]Please see Section 2 for information on how to exclude service classes from analysis.